Genetic risk scores do not improve asthma prediction in childhood

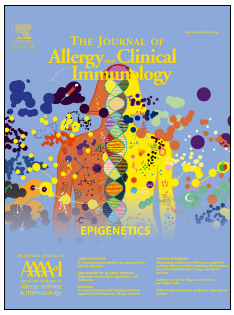
Dijk, F Nicole; Folkersma, Charlotte; Gruzieva, Olena; Kumar, Ashish; Wijga, Alet H; Gehring, Ulrike; Kull, Inger; Postma, Dirkje S; Vonk, Judith M; Melén, Erik

Published in:
Journal of Allergy and Clinical Immunology

DOI:
10.1016/j.jaci.2019.05.017

Publication date:
2019

Accepted Manuscript



Genetic risk scores do not improve asthma prediction in childhood

F. Nicole Dijk, MD, PhD, Charlotte Folkersma, Olena Gruzieva, MD, PhD, Ashish Kumar, PhD, Alet H. Wijga, PhD, Ulrike Gehring, PhD, Inger Kull, PhD, Dirkje S. Postma, MD, PhD, Judith M. Vonk, PhD, Erik Melén, MD, PhD, Gerard H. Koppelman, MD, PhD

PII: S0091-6749(19)30687-6

DOI: https://doi.org/10.1016/j.jaci.2019.05.017

Reference: YMAI 14014

To appear in: *Journal of Allergy and Clinical Immunology*

Received Date: 12 July 2018

Revised Date: 8 May 2019

Accepted Date: 14 May 2019

Please cite this article as: Dijk FN, Folkersma C, Gruzieva O, Kumar A, Wijga AH, Gehring U, Kull I, Postma DS, Vonk JM, Melén E, Koppelman GH, Genetic risk scores do not improve asthma prediction in childhood, *Journal of Allergy and Clinical Immunology* (2019), doi: https://doi.org/10.1016/j.jaci.2019.05.017.

Genetic risk scores do not improve asthma prediction in childhood.

F. Nicole Dijk, MD, PhD,[a,b] Charlotte Folkersma,[a,b] Olena Gruzieva, MD, PhD,[c,d] Ashish Kumar, PhD,[c,e,f] Alet H. Wijga, PhD,[g] Ulrike Gehring, PhD,[h] Inger Kull, PhD,[i,j] Dirkje S. Postma, MD, PhD,[k] Judith M. Vonk, PhD,[b,l] Erik Melén, MD, PhD,[c,i] and Gerard H. Koppelman, MD, PhD[a,b]

[a]University of Groningen, University Medical Center Groningen, Department of Pediatric Pulmonology and Pediatric Allergology, Beatrix Children's Hospital, Groningen, the Netherlands.

[b]University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD (GRIAC), Groningen, the Netherlands.

[c]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

[d]Centre for Occupational and Environmental Medicine, Stockholm County Council, Stockholm, Sweden

[e]Department of Public Health Epidemiology, Swiss Tropical and Public Health Institute, Basel, Switzerland

[f]University of Basel, Basel, Switzerland.

[g]Division of Environmental Epidemiology, Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands

[h]Center for Nutrition, Prevention, and Health Services, National Institute for Public Health and the Environment, Bilthoven, The Netherlands

[i]Sachs' Children and Youth Hospital, Södersjukhuset, Stockholm, Sweden

[j]Department of Clinical Science and Education, Södersjukhuset and Karolinska Institutet, Stockholm, Sweden

[k]University of Groningen, University Medical Center Groningen, Department of Pulmonology, Groningen, The Netherlands.

[I]University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, The Netherlands.

Corresponding author:

Gerard H. Koppelman, MD, PhD

Pediatric Pulmonology and Pediatric Allergology

Beatrix Children's Hospital

University Medical Center Groningen

PO Box 30.001

9700 RB Groningen, the Netherlands

Phone: + 31 50 3611036

Fax: + 31 50 3614235

E-mail: g.h.koppelman@umcg.nl

Declaration of all sources of funding :

This study was supported by the Netherlands Lung Foundation (grant no. 3.2.09.081JU); The Stichting Astma Bestrijding and the Ubbo Emmius Foundation. The BAMSE project was supported by The Swedish Research Council, The Swedish Heart-Lung Foundation, Stockholm County Council (ALF project) and the European Commission (The GABRIEL project; 018996).

Capsule summary:

Genetic risk scores have no added value above familial, perinatal and environmental risk factors in the first year of life to predict childhood asthma.

Key words:

asthma, prediction, genetic risk score, children

Disclosure of potential conflict of interest:

D.S. Postma declares that the University of Groningen has received money for D.S. Postma regarding a grant for research from Astra Zeneca, Chiesi, Genentec, GSK and Roche. Fees for consultancies were given to the University of Groningen by Astra Zeneca, Chiesi, and GSK. G.K. Koppelman received grants from the Lung Foundation of the Netherlands, BBMRI-NL, the UBBO emmius Foundation, during the conduct of the study; and he received grants from Lung Foundation of the Netherlands, GSK, Tetri Foundation, Vertex, TEVA the Netherlands, outside the submitted work. The rest of the authors declare that they have no relevant conflict of interests.

To the Editor,

Thirty to 50% of preschool children experience asthma-like symptoms, such as wheezing,[1–3] but only approximately 30% of these children will develop asthma. Due to the non-specific symptoms of asthma at preschool age and the lack of a diagnostic test for asthma in this age group, it is difficult to determine which child will develop asthma. Several prediction models based on family, personal and environmental factors have been developed to improve the early diagnosis of asthma,[2,3] yet these are of modest clinical value.[4] In addition, these models are based on children with respiratory symptoms, while asthma prediction at a time point when no clinical symptoms have occurred may identify children at risk for asthma to start early preventative measures.

It has been proposed that genetics may improve asthma prediction.[4] Recently, two consortia published the results on large meta-analyses of genome wide association studies (GWAS), which doubled the number of genetic variants that are associated with asthma.[5,6] The Trans-National Asthma Genetic Consortium (TAGC) consortium described 18 loci to be associated with asthma in a multi-ancestry meta-analysis in 142,000 subjects,[5] with 5 additional loci specifically related to pediatric asthma. Moreover, the SHARE consortium discovered 136 independent genetic variants to be associated with allergic disease (asthma, hay fever or eczema) in 360,000 subjects, with almost all variants contributing to either disease.[6] These asthma associated variants from TAGC and SHARE offer the opportunity to investigate asthma prediction based on genetic risk scores.

We generated a prediction model for asthma in the first 8 years of life based on the combination of family, perinatal, environmental and genetic risk factors, with the aim to investigate the added value of genetics at predicting childhood asthma determined by easy available factors known in the first year of life. Asthma definition was based on asthma ever from age 3 till age 8 years, in which cases had one or more of the following three criteria; (1) having ≥1 more attacks of wheeze in the last 12 months, (2) ≥1 events of shortness of breath

(dyspnea) in the last 12 months, or (3) prescription of inhalation steroids for respiratory or lung problems prescribed by a doctor in last 12 months.

We used data from the Prevalence and Incidence of Asthma and Mite Allergy birth cohort (PIAMA)[7] with inclusion of 1,968 children (see Online Repository). With univariate and multivariate logistic regression analysis, familial, perinatal and environmental risk scores were made based on variables which previously predicted asthma in children experiencing respiratory symptoms.[2] We selected independent single nucleotide polymorphisms (SNPs) and calculated weighted genetic risk score (GRS) based on the TAGC and SHARE data (see Online Repository).[5,6] Receiver-operating characteristics (ROC) analysis was performed to test the added value of the GRSs to the familial, perinatal and environmental scores. Since the predictors of the non-genetic risk scores were generated from PIAMA data we tested for optimism caused by overfitting with the use of an internal bootstrap validation approach (using the R package 'rms'). The predicted probabilities of the separate risk scores were categorized into deciles to analyze the discriminative performance of each score. Replication of the models obtained in PIAMA was performed in BAMSE (Children/Barn, Allergy, Milieu, Stockholm, Epidemiology) (n=427), a Swedish birth cohort, with a comparable design to PIAMA.[8]

Of the 1,968 children with genotype data in our study, 1,858 children had information on the presence of asthma in the first 8 years of life. Of these, 42.6% (n=792) had asthma ever in the first 8 years of life (Table I).

The combined familial, perinatal and environmental risk score included parental allergy (OR (95% CI)=1.38 (1.08-1.76), parents allergic to pets (1.43 (1.10-1.85), parental inhaled medication (1.54 (1.17-2.04), siblings with asthma (2.46 (1.52-3.99), low parental education (1.33 (1.07-1.65), male gender (1.44 (1.20-1.74), breastfeeding <16 weeks (1.32 (1.09-1.60), low birth weight <2500g (2.15 (1.24-3.70), pets at home during pregnancy (1.21 (1.00-1.46), smoking mother during pregnancy (1.45 (1.13-1.88), and older siblings living at home (1.20 (1.00-1.45) (Table E1). Association analyses with asthma separately for familial, perinatal,

environmental and genetic risk score indicated that the familial risk score had the strongest prediction (PIAMA; OR=1.25, P =3.17x10^{-19}, BAMSE; OR=1.46, P =3.17x10^{-13}) (Table E2). The combined model of familial, perinatal and environmental factors showed moderate discrimination (area under receiver operating characteristic curve (AUC)=0.65), with similar predictive properties of this model in the BAMSE study (AUC=0.67). Optimism corrected AUC in PIAMA was 0.65 which indicates no overfitting of data. In PIAMA , including the TAGC GRS in the models with the familial, perinatal and environmental risk score showed an AUC of 0.66, whereas the inclusion of the SHARE GRS had an AUC of 0.65 (Fig 1A-B). There was no improvement over the risk prediction based on familial, perinatal and environmental factors (AUC difference between familial, perinatal and environmental factors solely and combined with GRS: TAGC; Z=-0.55, P=0.29, SHARE; Z-0.0, P=0.5) Replication analyses in BAMSE showed similar results with AUC of 0.69 (AUC difference between familial, perinatal and environmental factors solely and combined with GRS: TAGC; Z=-0.55, P=0.29, SHARE; Z-0.83, P=0.2) (Fig 1C-D). Discriminative analysis showed best predictive probability for the familial risk score (Fig E1A-B, E2A-B). In PIAMA the results did not change when we used a more specific asthma diagnosis as the outcome, doctor's diagnosed asthma at age 8, which will exclude the transient wheezers (model of familial, perinatal and environmental factors (AUC=0.64) combined with TAGC GRS: AUC=0.64; SHARE GRS: AUC=0.64). It has been suggested that genetic risk prediction may improve when adding additional SNPs that are associated with the disease, albeit not at genome wide significance threshold. To investigate this possibility, we performed additional predictive analysis of the genetic risk scores from the SHARE study in PIAMA using more liberal P value thresholds of 1x10^{-6} and 1x10^{-4}. However, this did not improve genetic risk prediction for asthma ever in the first 8 years of life with combined familial, perinatal, environmental and GRS AUC values of 0.65 and 0.65, respectively (see Online Repository).

Identifying children at high risk for asthma development is important for prevention and installation of early treatment. However, the GRSs based on SNPs from the largest asthma GWAS did not improve asthma prediction over familial, perinatal and environmental factors. Asthma in childhood is a highly heterogeneous disease, with different genes being related to different sub-types of asthma. The fact that we used a more common asthma definition with the selection of all children with respiratory symptoms in the first 8 years of life and no selection on disease specificity, could have influenced our results, although using a more strict definition (i.e. doctors diagnosed asthma) led to the same conclusion. For the calculation of the TAGC GRS we therefore added loci specifically related to pediatric asthma. In addition to this, to improve the prediction of asthma sub-types an even more specific (subtype-related) selection of SNPs could be beneficial for generating GRSs.

Asthma has a strong genetic contribution. However, based on the most recent insights in asthma genetics, genetic variants have no added value in predicting non-specific asthma. This can be explained in several ways. The known heritability of asthma is due to common SNPs of modest effect, resulting in many children carrying risk alleles but not having asthma. Second, although the number of risk SNPs has increased considerably in the past years, these SNPs still explain only a small fraction of asthma heritability. We also acknowledge that a substantially larger study may have yielded a significant, but small, increase in AUC values after inclusion of the genetic risk score. We show in our paper that variation in *P* value thresholds for GRS SNP selection made no difference in asthma prediction. This is underlined by the findings by Zhang et al.[9], who propose subsequently to focus more on effect size distributions than *P* values for selection of SNPs for disease prediction. These novel methodological approaches may benefit future genetic risk prediction in asthma. Better prediction may also depend on our ability to define different sub-types of asthma with shared etiology. Moreover, better modeling of potential interactions between genes and environmental factors[E1] may be needed to accurately predict asthma in future studies.

Authors:

F. Nicole Dijk, MD, PhD,[a,b] Charlotte Folkersma,[a,b] Olena Gruzieva, MD, PhD,[c,d] Ashish Kumar, PhD,[c,e,f] Alet H. Wijga, PhD,[g] Ulrike Gehring, PhD,[h] Inger Kull, PhD,[i,j] Dirkje S. Postma, MD, PhD,[k] Judith M. Vonk, PhD,[b,l] Erik Melén, MD, PhD,[c,i] and Gerard H. Koppelman, MD, PhD[a,b]

Authors affiliations:

[a]University of Groningen, University Medical Center Groningen, Department of Pediatric Pulmonology and Pediatric Allergology, Beatrix Children's Hospital, Groningen, the Netherlands.

[b]University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD (GRIAC), Groningen, the Netherlands.

[c]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

[d]Centre for Occupational and Environmental Medicine, Stockholm County Council, Stockholm, Sweden

[e]Department of Public Health Epidemiology, Swiss Tropical and Public Health Institute, Basel, Switzerland

[f]University of Basel, Basel, Switzerland.

[g]Division of Environmental Epidemiology, Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands

[h]Center for Nutrition, Prevention, and Health Services, National Institute for Public Health and the Environment, Bilthoven, The Netherlands

[i]Sachs' Children and Youth Hospital, Södersjukhuset, Stockholm, Sweden

[j]Department of Clinical Science and Education, Södersjukhuset and Karolinska Institutet, Stockholm, Sweden

[k]University of Groningen, University Medical Center Groningen, Department of Pulmonology, Groningen, The Netherlands.

[l]University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, The Netherlands.

References

1. Sears MR. Predicting asthma outcomes. J Allergy Clin Immunol 2015;136:829–36.
2. Caudri D, Wijga A, A Schipper CM, Hoekstra M, Postma DS, Koppelman GH, et al. Predicting the long-term prognosis of children with symptoms suggestive of asthma at preschool age. J Allergy Clin Immunol 2009;124:903–10.
3. Hafkamp-De Groen E, Lingsma HF, Caudri D, Levie D, Wijga A, Koppelman GH, et al. Predicting asthma in preschool children with asthma-like symptoms: Validating and updating the PIAMA risk score. J Allergy Clin Immunol 2013;132:1303-10
4. Colicino S, Munblit D, Minelli C, Custovic A, Cullinan P. Validation of childhood asthma predictive tools: A systematic review. Clin Exp Allergy. 2019;doi:10.1111/cea.13336.
5. Demenais F, Margaritte-Jeannin P, Barnes KC, Cookson WOC, Altmüller J, Ang W, et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. Nat Genet 2018;50:42–50.
6. Ferreira MA, Vonk JM, Baurecht H, Marenholz I, Tian C, Hoffman JD, et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. Nat Genet. 2017;49:1752–7.
7. Wijga AH, Kerkhof M, Gehring U, de Jongste JC, Postma DS, Aalberse RC, et al. Cohort profile: the prevention and incidence of asthma and mite allergy (PIAMA) birth cohort. Int J Epidemiol. 2014;43:527–35.
8. Kull I, Melen E, Alm J, Hallberg J, Svartengren M, van Hage M, et al. Breast-feeding in relation to asthma, lung function, and sensitization in young schoolchildren. J Allergy Clin Immunol. 2010;125:1013–9.
9. Zhang Y, Qi G, Park JH, Chatterjee N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. Nat Genet. 2018;50:1318-26.

Figure Legend

FIG 1. Receiver-operating characteristic graphs of familial, perinatal and environmental risk score with and without the combination of genetic risk scores in PIAMA and BAMSE. **(A-B)** Contains the familial, perinatal and environmental risk score in the PIAMA birth cohort combined with the genetic risk score (GRS) from the TAGC consortium **(A)**,and the GRS from SHARE consortium **(B)**. **(C-D)** Contains the familial, perinatal and environmental risk score in the BAMSE birth cohort combined with the GRS from the TAGC consortium **(C)**, and the GRS from SHARE consortium **(D)**.

238 **TABLE I. General characteristics of the study population and univariate analysis on asthma ever at age 8 years**
239

Candidate predictor	PIAMA			BAMSE		
	All children (n = 1858)	OR (95% CI)	*P* value	All children (n = 427)	OR (95% CI)	*P* value
Familial factors % (n total)						
Parental allergy†	55.5 (1858)	1.87 (1.55-2.26)	6.50×10^{-11}	28.6 (423)	2.74 (1.76-4.28)	8.32×10^{-6}
Parental asthma	16.1 (1848)	1.88 (1.47-2.42)	7.35×10^{-7}	26.2 (423)	5.01 (3.04-8.26)	2.52×10^{-10}
Parental allergy house dust (mite)	34.0 (1835)	1.77 (1.46-2.15)	9.93×10^{-9}	NA	NA	NA
Parental allergy to pets	30.6 (1836)	1.91 (1.56-2.33)	2.83×10^{-10}	37.1 (423)	2.83 (1.87-4.26)	7.36×10^{-7}
Parental hay fever	41.8 (1832)	1.51 (1.25-1.82)	1.9×10^{-5}	47.5 (423)	2.41 (1.63-3.56)	0.00001
Parental inhaled medication	16.2 (1835)	2.20 (1.71-2.83)	9.72×10^{-10}	30.5 (423)	4.65 (2.93-7.39)	7.64×10^{-11}
Low parental education‡	26.7 (1840)	1.25 (1.02-1.54)	0.034	51.1(427)	1.89 (1.28-2.77)	0.001
Allergic siblings§	20.5 (1845)	1.52 (1.21-1.90)	3.12×10^{-4}	10.4 (222)	2.45 (0.97-6.21)	0.06
Asthma siblings	4.6 (1856)	2.88 (1.82-4.58)	7×10^{-6}	6.6 (427)	3.91 (1.55-9.86)	0.004
Eczema siblings	17.8 (1846)	1.36 (1.07-1.73)	0.012	20.2 (223)	1.02 (0.53-1.97)	0.95
Hay fever siblings	1.98 (1851)	2.11 (1.04-4.26)	0.038	8.1 (223)	0.97 (0.37-2.55)	0.95
Perinatal factors % (n total)						
Male gender	51.1 (1858)	1.46 (1.21-1.75)	6.9×10^{-5}	56.2 (427)	2.18 (1.47-3.21)	0.0001
Low birth weight <2500g	3.1 (1855)	2.20 (1.28-3.78)	0.004	2.4 (424)	4.04 (0.85-19.3)	0.08
Any breastfeeding	84.4 (1849)	1.07 (0.83-1.37)	0.619	98.3 (424)	3.06 (0.61-15.3)	0.174
Breastfeeding <16 weeks	63.4 (1849)	1.35 (1.12-1.64)	0.002	18.4 (423)	1.77 (1.07-2.92)	0.03
Delivery:						
Term (≥37-≤42 weeks)	92.0 (1855)	Ref	Ref	89.2 (427)	Ref	Ref
Preterm (<37 weeks)	4.7 (1855)	1.55 (1.01-2.39)	0.047	7.0 (427)	2.05 (0.94-4.50)	0.073
Postterm (>42 weeks)	3.3 (1855)	1.17 (0.70-1.96)	0.545	3.8 (427)	0.80 (0.29-2.19)	0.662
Born by caeserian section	8.5 (1841)	1.07 (0.77-1.49)	0.696	16.7 (427)	1.43 (0.85-2.39)	0.174
Environmental factors % (n total)						
Pets at home during pregnancy	44.4 (1856)	1.22 (1.01-1.47)	0.036	12.7 (427)	0.76 (0.43-1.35)	0.35
Smoking mother during pregnancy‖	15.5 (1842)	1.50 (1.16-1.93)	0.002	11.0 (427)	2.30 (1.20-4.38)	0.01

Older siblings living in home	51.7 (1858)	1.18 (0.98-1.41)	0.085	52.2 (424)	1.03 (0.70-150)	0.89

OR, Odds ratio; CI, Confidence interval.

* In PIAMA asthma is defined as having one or more of the following three criteria; (1) having ≥1 more attacks of wheeze in the last 12 months, (2) ≥1 events of shortness of breath (dyspnea) in the last 12 months, (3) prescription of inhalation steroids for respiratory or lung problems prescribed by a doctor in last 12 months. Since no data on shortness of breath was available in BAMSE we used an adjusted asthma definition in which asthma was defined as having one or more of the following two criteria; (1) having ≥1 more attacks of wheeze in the last 12 months, (2) prescription of inhalation steroids for respiratory or lung problems prescribed by a doctor in last 12 months.

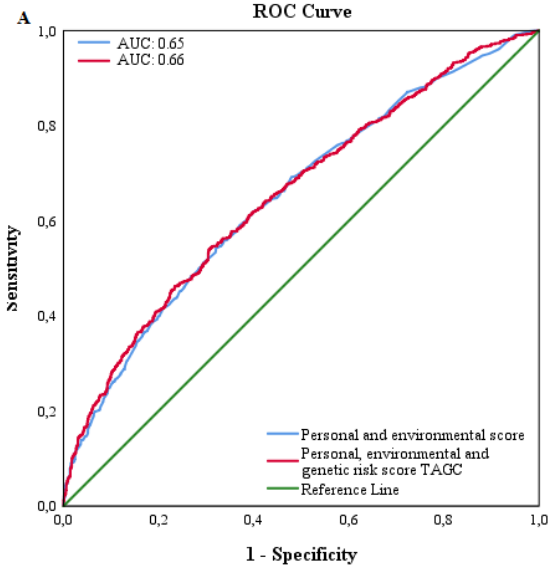
† In PIAMA parental allergy is based on parental asthma ever and/or current house dust (mite) allergy and/or pet allergy and/or hay fever. In BAMSE it is based on mother and/or father with doctor's diagnosis of asthma and/or doctor's diagnosis of hay fever in combination with pollen allergy at baseline.

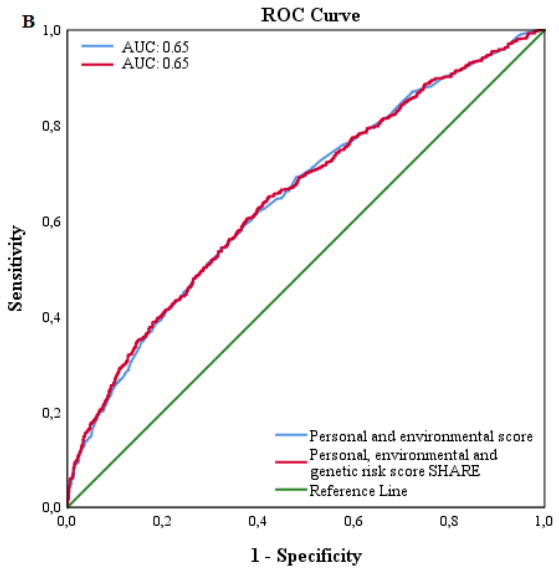
‡ In PIAMA parental education is defined as an education less than the level of a bachelor's/master's degree (HBO/University in Dutch system) for at least 1 of the parents. In BAMSE it is defined as an education level less than university grade for both of the parents.

§ In PIAMA a sibling with allergy is based on a sibling with asthma ever and/or eczema and/or hay fever. In BAMSE it is based on allergy to furred animals or pollen.

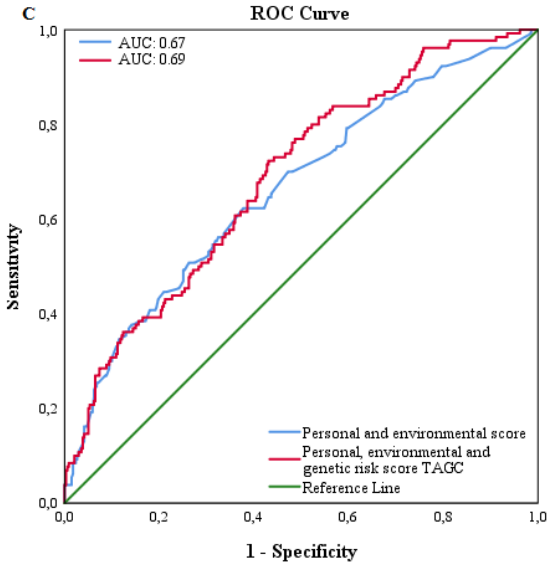
ǁ In PIAMA smoking during pregnancy is defined as smoking at least the first 4 weeks of pregnancy. In BAMSE it is defined as smoking at least one cigarette per day in any point of time during pregnancy.

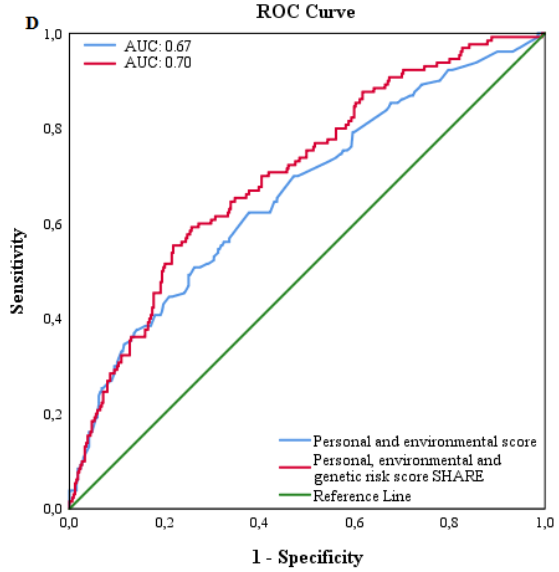
PIAMA birth cohort





BAMSE birth cohort





Online Repository

Genetic risk scores do not improve asthma prediction in childhood

F. Nicole Dijk, Charlotte Folkersma, Olena Gruzieva, Ashish Kumar, Alet H. Wijga, Ulrike Gehring, Inger Kull, Dirkje S. Postma, Judith M. Vonk, Erik Melén, and Gerard H. Koppelman

Supplemental Methods

Study populations

PIAMA cohort

The PIAMA study is a multicenter birth cohort, which was initiated in 1996. 7862 women (2779 with allergy and 5083 without allergy) were invited to participate in the study; 3963 live-born children participated the study (1327 with a mother with allergy were defined as high-risk, and 2726 children with a mother without allergy were defined as low-risk). Questionnaires for parental completion, partly based on the International Study of Asthma and Allergies in Childhood core questionnaires, were sent to the parents during pregnancy, when the children were aged 3 and 12 months, yearly thereafter up to the age of 8 years, at the age of 11 years, 14 years 16 years, and 17 years. All 1327 high-risk children and a random sample of 663 low-risk children were selected for an extensive medical examination at age 4 and 8 years. Blood or a buccal brush was used for DNA extraction during the extensive medical examination group at age 4 and in all children at age 8.

At age 8 years, 92% of the baseline population was still in the study, and therefore our study focused on the first 8 years of life. Combined phenotypic and genotype data for this study was available for 1,968 children.

The study protocol was approved by the Medical Ethical Committees of the participating university hospitals and all participants gave written, informed consent. A detailed description of the cohort outline has been published previously.[E2]

BAMSE

Between 1994 and 1996, 4,089 newborn infants were recruited in the BAMSE (Children/Barn, Allergy, Milieu, Stockholm, Epidemiology) study, and questionnaire data on baseline study characteristics were obtained.[E3] The recruitment area included central and north-western parts of Stockholm. At approximately one, two, four, and eight years of age, parents completed questionnaires on their children's symptoms related to asthma and other allergic diseases. The response rates were 96%, 94%, 92% and 84%, respectively.

At ages 4, and 8 years, blood samples were collected in 2,605 (63.7%), and 2,470 (60.4%) children, respectively.

DNA was extracted from 2,033 samples at 8 years after exclusion of samples with too little blood, lack of questionnaire data, or if parental consent to genetic analysis of the sample was not obtained. From these samples, all children with a doctor's diagnosis of asthma ever were selected as cases (n=273) and a random sample of children with no history of asthma or other allergic diseases was selected as controls (n=273). After Quality Control (QC) a total of 239 cases and 246 controls, all of Caucasian ancestry, were retained for genetic analyses.[E4]

Genotyping and imputation

PIAMA cohort

Children from the PIAMA cohort were genotyped on three different platforms. 1377 children were genotyped with the Illumina Omni Express Exome (OEE) chip, whereas 288 children were genotyped with the Illumina Omni Express (OE) chip (Illumina Inc, San Diego, CA), both with the use of an Illumina BeadArray Reader and Iscan at the Genomics Facility of the University Medical Center Groningen, Groningen, The Netherlands. DNA of 404 children was genotyped with the Illumina Human610 (HM610) quad array and the use of the Ilumina Beadarray reader and Iscans at the Centre National de Génotypage (CNG, Evry, France) as part of the GABRIEL consortium.[E5]

Quality control inclusion measures per chip on the individuals included a missing genotype call rate <0.03, IBS <0.1875 and a heterozygosity rate deviating <4SD from the mean. Males with >1% heterozygote SNPs on chromosome X were excluded. Ethnicity was assessed using principal component analyses with HapMap CEU, CHB+JPT, and YRI reference panels, only Caucasians subject were included.[E6]

QC measures per SNP included missing genotype call rate <0.05, MAF >0.05 and Hardy-Weinberg equilibrium *P*-value >10-6. SNPs being >1% heterozygous in males on chromosome X were excluded.

Base pair positions of SNPs on the HM610 chip were converted to genome build 37, in accordance with the OEE chip and the OE chip.

The strand was determined of each SNP and on the different platforms, and if necessary converted to the positive strand. SNPs with unknown strand orientation were removed. Discordant genotypes of duplicate SNPs were set to missing. SNPs that showed large differences in allele frequencies between platforms (>15 %) were either recoded (i.e. alleles were swapped) in case of an A/T or C/G SNP (and rechecked) or removed in other cases. Duplicate individuals between the platforms were considered sampling errors and both individuals were removed.

The single chips were matched to the 1000G reference set with respect to basepair positions Resemblance between the chip and the 1000G European panel (EUR) of rs-numbers, alleles, and allele frequencies of SNPs on the autosomal chromosomes were checked and if discrepant deleted.

After quality control, a total of 1968 individuals remained, with the presence of 873 (44,4%) high-risk children, Imputation was performed per platform using IMPUTE 2.0.[E7] against the reference data set of the ALL panel of 1000G (version 3, March 2012).[E8] After imputation, only SNPs of high quality (info-score IMPUTE ≥ 0.7) were selected per chip. We removed SNPs that showed discrepancy between chips in allele frequency (> 15 %) (N=1795). Rs-numbers and insertions or deletions were separately merged using GTOOL (http://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html) due to potential localization at the same base-pair position. The obtained files were combined into one dataset (SNPs N=11,713,219) that was used for further analyses.

BAMSE

Genotyping was done on the Illumina Human610 Quad platform at the Centre National de Génotypage in Evry, France under the GABRIEL project framework.[E5]

For imputation, the genotyped SNPs were filtered at - call rate >95%, Hardy Weinberg *P*-value > $1x10^{-6}$ and MAF > 0.01; and sample call rate > 95%; and 515,445 SNPs remained after quality control. These were imputed using MiniMac release stamp 2012-11-16 and the GIANT ALL reference panel, phase 1 v3.20101123 onto N=30,061,897 variants. The resultant SNPs were filtered for imputation quality threshold at Rsq >=0.3.

Outcome variable

PIAMA

The primary outcome variable of this study is based on asthma ever at age 8 years, in which asthma is defined by the following characteristics: one or more attacks of wheeze in the last 12 months, or one or more events of shortness of breath (dyspnea) in the last 12 months, or prescription of inhaled corticosteroids for respiratory or lung problems prescribed by a doctor in the last 12 months. A child who had one or more of these characteristics was categorized as having 'asthma'. A child who had none of these characteristics was categorized as 'not having asthma'. At 1 and 2 years of age, data on shortness of breath is not available and the data on steroids use is limited. Therefore our outcome variable is based on asthma ever from age 3 till age 8 years.

We acknowledge that with our asthma ever definition we select all children with respiratory symptoms in the first 8 years of live and that some included children will not develop asthma but have respiratory symptoms. Therefore we performed our analyses as well on a the PIAMA variable doctors diagnosis of asthma at age 8, which is defined as asthma diagnosed ever by a doctor and asthma in the last 12 months at the age of 8 years.

BAMSE

In BAMSE no data were available on one or more events of shortness of breath (dyspnea) in the last 12 months. Therefore, we used an adjusted diagnosis of asthma ever at age 8 year based on: (1) one or more attacks of wheeze in the last 12 months, (2) prescription of inhaled corticosteroids for respiratory or lung problems prescribed by a doctor in the last 12 months.

A child who had at least one of these characteristics was categorized as having 'asthma'. A child who had none of these characteristics was categorized as 'not having asthma'.

Familial, perinatal and environmental variables

PIAMA

In the PIAMA cohort a prediction model for asthma at age 8 years in preschool children who have asthma-like symptoms was previously published.[E9,E10] We used these data, in combination with other asthma associated studies performed in PIAMA, to select candidate predictor variables present at birth.[E9-E11] We took into account that for a prediction model the variables have to be available in the first months of life and not involve invasive tests. We divided the candidate predictors in three groups; (1) familial, (2) perinatal and (3) environmental factors. If the candidate predictors were not present in our dataset we searched for a surrogate or combined marker.

BAMSE

In the BAMSE cohort, predictor variables were defined in accordance or as similar as possible with PIAMA definitions (as described above) using the data collected via parental questionnaires as well as Medical Birth Registry.

Reference datasets and SNP selection

TAGC/SHARE consortia

We used the findings from the two largest asthma and allergy genetics consortia to summarize different single nucleotide polymorphisms (SNPs) associated with asthma. The first study is the Trans-National Asthma Genetic Consortium (TAGC)[E12] in which the largest meta-analysis of asthma GWAS (23,948 cases, 118,538 controls) was conducted from multi-ancestry populations. New asthma loci were identified and associations at known asthma loci were confirmed. For our analysis we selected the 18 lead SNPs and 5 specific variants which

were associated with pediatric asthma. Of those selected SNPs, 1 was missing leaving 22 SNPS to be used for further analyses.

The second study is the SHARE Consortium[E13], a large study about the three most common atopic diseases, asthma, hay fever (allergic rhinitis) and eczema (atopic dermatitis). A GWAS (360,838 samples) was performed on an allergic disease phenotype. Because these diseases frequently occur in the same individuals and partly have a shared genetic origin, they identified individual genetic risk variants shared between asthma, hay fever and eczema. They also identified 6 variants that had stronger effect in one allergic disease, which confirmed that the majority acted as shared risk factors. The main association result showed 136 SNPs independently associated with risk of allergic disease, of which 133 were present in PIAMA and used for this study.

We performed additional sensitivity analyses using results of the SHARE consortium dataset to investigate if including SNPs that were less significantly associated would improve asthma prediction. To this aim, we downloaded the results from the SHARE consortium; and applied two less stringent significance cut-offs: $1x10^{-6}$ and $1x10^{-4}$. Results of one of the largest cohorts, 23andme, were not included in this download and therefor these results differ slightly from the main results of the SHARE-analysis.

To select independently associated SNPs based on the selected p-value thresholds we used Genome Wide Complex Trait (GCTA) software (https://cnsgenomics.com/software/gcta/). We used data from the LifeLines cohort study as a reference set.[E14] LifeLines is a large Dutch population-based cohort study and biobank that was established as a resource for research on complex interactions between environmental, phenotypic and genomic factors in the development of chronic diseases and healthy ageing. Genomic imputed data, based on the 1000G reference set, are available of 13,386 participants. With the use of the --cojo-slct option independent SNPs were selected using the default settings of the program (i.e. –cojo-wind 10000 and --cojo-collinear=0.9). Weighted GRSs were calculated in the same manner as for the TAGC and SHARE GRS already stated in the paper. This resulted in a SNP set of 175 SNPs for the $1x10^{-6}$ threshold and 634 SNPs for the $1x10^{-4}$ threshold.

Statistical analysis

Predictive modeling of familial, personal and environmental factors.

We performed univariate logistic regression to assess the predictive value of the candidate predictors on asthma. We selected predictors of each category (familial factors, perinatal factors, environmental factors) which were used in the previous asthma prediction model for children with symptoms in the PIAMA birth cohort[E9,E10] and had a *P*-value <0.10 in our univariate analysis. All these variables were entered in a multivariate logistic regression model. Using a stepwise backward regression strategy we selected our final predictor models (one model per category) addressing the change in P values and Cox-Snell and Nagelkerke R-square (closest to 1) to select the best model.

Familial, perinatal and environmental risk score

To develop these risk scores we created a weighted score per category using the regression coefficients from the final multivariate models to determine the score for each variable. In the weighted score per category the scores were calculated and rescored in a range from 0-10 giving equal weight to each category. With this score the variables were tested in a total model instead of the separate categories (familial, perinatal and environmental factors). The weight for each variable was calculated by using the regression coefficients from the model including all predictor variables.

Association analyses with asthma ever at age 8 years were performed per familial, perinatal and environmental risk score (see table 1 and E1 for selected variables).

Genetic risk score

We generated weighted genetic risk scores (GRSs) based on significant SNPs selected from the previously published TAGC ($P < 5x10^{-8}$) and SHARE ($P < 3x10^{-8}$) studies and used the reported ORs of the meta-analysis for weighing the GRSs.[E12,E13] The GRSs were calculated with the use of SPSS for Windows, Version 24.0 (IBM SPSS Statistics). The dosages of the asthma risk alleles were calculated and were summed up to develop the unweighted GRS.

To construct a weighted GRS we took into account the effect sizes of the SNPs. For developing the weighted score we used the odds ratios and multiplied them with the dosages of the risk alleles, summed them up and then divided the results by two times the sum of the weights.

To calculate the TAGC and SHARE GRS in BAMSE, data was available for 19 of the 23 and 133 of the 136 SNPs, respectively. For the additional sensitivity analysis in SHARE using less stringent significant thresholds, data was available for 172 and 613 SNPs, respectively. Association analyses with asthma ever at age 8 years were performed per genetic risk score (threshold: TAGC; $P < 5x10^{-8}$ and SHARE; $P < 3x10^{-8}$).

Combination of risk scores

We combined the familial, perinatal and environmental risk scores with the GRSs to construct two final models; (1) the familial, perinatal and environmental risk scores with the TAGC GRS, and (2) the familial, perinatal and environmental risk scores with the SHARE GRS. To determine the discriminative ability we conducted a receiver-operating characteristic (ROC) curve. ROC-curves of the risk models were made using the predicted probabilities from the logistic regression models for the asthma ever at age 8 years definition. The area under the ROC-curve (AUC) was calculated. The AUC ranges from 0 to 1, in which a value of 0.5 means not better than chance, and a value closer to 1 means a better discrimination.[E15] Significant differences between the AUCs were tested.[E16] The predicted probabilities of the final models were categorized into deciles. For each decile of predicted asthma risk the mean of the 4 included risk scores were plotted.

Validation of the model

Optimism caused by overfitting was investigated by internal validation of the familial, perinatal and environmental weighted risk scores in PIAMA with the use of the bootstrapping methods implemented in the R-package rms. The number of bootstraps samples was set to 1000.

Supplemental Results

General characteristics and univariate analysis

General characteristics of the study population and univariate analysis on asthma ever at age 8 years in PIAMA and BAMSE are shown in Table 1.

Of the 1,968 children with genotype data in the PIAMA cohort, 1,858 children had information on the presence of asthma in the first 8 years of life. Data on doctor's diagnoses at age 8 was available for 1,794 children. In BAMSE, 427 of the 485 children with genetic data had information on the presence of asthma ever at age 8 year and were selected for this study. In the PIAMA study population with data on asthma ever at age 8 years the percentage of high risk children (defined as children of allergic mothers) was 36.9% as compared to 31.2% in the PIAMA population as a whole and 29% in the general population of pregnant women from which PIAMA participants were recruited.

The prevalence of children with allergic parents is 55.5% in PIAMA, compared to 28.6% in BAMSE. However, the number of children with asthma ever at age 8 were comparable between the cohorts, with 42.6% (N=792) cases in PIAMA and 50.4% (N=215) cases present in BAMSE. In PIAMA 4.0% (N=71) children had doctors diagnosis at age 8.

Multivariate analysis and calculation of familial, perinatal and environmental risk scores

Variables that were found to have a significant association with the risk of asthma development were further tested in the multivariable analysis. Despite being significantly associated with asthma in the univariate analysis, parental allergy to house dust (mite) (P = 0.7) was not significantly associated in the multivariate analysis, nor was there a significant association between asthma and having siblings with eczema (P = 0.2) or hay fever (P = 0.6). The final model of familial risk score contained the variables parental allergy, parents allergic to pets, low parental education, parental inhaled medication to improve breathing, and siblings with asthma. Regarding the perinatal factors, both preterm (P = 0.6) and post-term (P = 0.6) delivery were not significant in the multivariate analysis and were removed

from the model. Male gender, low birth weight <2500 g, and breastfeeding <16 weeks were the final variables for the perinatal risk score. Selected variables for the environmental risk score were pets at home during pregnancy, smoking mother during pregnancy, and older siblings living at home. A risk score was calculated by using the regression coefficient of each predictor variable shown in Table 1. A weighted risk score per category was developed by assigning points for each variable based on the regression coefficient with a range from 0 to 10. Scores per category are shown in Table E1.

Combined risk scores

Calibration of our models showed that the mean familial risk score increased most per decile, in contrast to the almost horizontal lines of the GRSs from TAGC and SHARE (Fig E1A-B). The increases per decile of the predicted probability of the perinatal and environmental scores were both slightly lower than the family score.

Internal validation of the model

The optimism corrected C-statistic (i.e. AUC) of the familial, perinatal and environmental weighted risk score in PIAMA showed a value of 0.65 (uncorrected AUC=0.65).

Replication results

ROC-curves of the two risk models in BAMSE; (1) the familial, perinatal and environmental risk scores with the TAGC GRS, and (2) the familial, perinatal and environmental risk scores with the SHARE GRS are shown in Fig 1C-D.

Fig E2A-B shows the observed mean score of the predicted probability for each decile. The horizontal lines of the TAGC and SHARE based GRSs indicate a low ability to predict asthma. Familial and perinatal scores show better prediction when compared to the genetic and environmental risk scores.

Supplemental References

E1. Bønnelykke K, Ober C. Leveraging gene-environment interactions and endotypes for asthma gene discovery. J Allergy Clin Immunol 2016;137:667–79.

E2. Wijga AH, Kerkhof M, Gehring U, de Jongste JC, Postma DS, Aalberse RC, et al. Cohort profile: the prevention and incidence of asthma and mite allergy (PIAMA) birth cohort. Int J Epidemiol 2014;43:527–35.

E3. Wickman M, Kull I, Pershagen G, Nordvall SL. The BAMSE project: presentation of a prospective longitudinal birth cohort study. Pediatr Allergy Immunol 2002;13 Suppl 1:11–3.

E4. Melén E, Granell R, Kogevinas M, Strachan D, Gonzalez JR, Wjst M, et al. Genome-wide association study of body mass index in 23 000 individuals with and without asthma. Clin Exp Allergy. 2013;43:463–74.

E5. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 2010;363:1211–21.

E6. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–75.

E7. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet 2012;44:955–9.

E8. Consortium 1000 Genomes Project, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56–65.

E9. Caudri D, Wijga A, A. Schipper CM, Hoekstra M, Postma DS, Koppelman GH, et al. Predicting the long-term prognosis of children with symptoms suggestive of asthma at preschool age. J Allergy Clin Immunol 2009;124:903–910.

E10. Hafkamp-De Groen E, Lingsma HF, Caudri D, Levie D, Wijga A, Koppelman GH, et al.

Predicting asthma in preschool children with asthma-like symptoms: Validating and updating the PIAMA risk score. J Allergy Clin Immunol 2013;132:1303-10.

E11. Roduit C, Scholtens S, de Jongste JC, Wijga AH, Gerritsen J, Postma DS, et al. Asthma at 8 years of age in children born by caesarean section. Thorax 2009;64:107–13.

E12. Demenais F, Margaritte-Jeannin P, Barnes KC, Cookson WOC, Altmüller J, Ang W, et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. Nat Genet 2018;50:42–50.

E13. Ferreira MA, Vonk JM, Baurecht H, Marenholz I, Tian C, Hoffman JD, et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. Nat Genet 2017;49:1752–7.

E14. Scholtens S, Dotinga A, Stolk RP, van Zon SK, van Dijk F, Wijmenga C, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. Int J Epidemiol. 2014;44:1172–80.

E15. Harrell FE. (2001). Regression Modelling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. New York: Springer.

E16. Hanley J, McNeil B. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. Radiol 1982;1982;29–36.

Supplementary Figure Legends

FIG E1. Mean predicted scores per decile for each score in the PIAMA cohort. **(A-B)** Contains the familial, perinatal and environmental risk score in the PIAMA birth cohort combined with the genetic risk score (GRS) from the TAGC consortium **(A)**, and the GRS from SHARE consortium **(B)**. Scores are all based on asthma ever at age 8 years, in which asthma is defined as having one or more of the following three criteria; (1) having ≥1 more attacks of wheeze in the last 12 months, (2) ≥1 events of shortness of breath (dyspnea) in the last 12 months, (3) prescription of inhalation steroids for respiratory or lung problems prescribed by a doctor in last 12 months.

FIG E2. Mean predicted scores per decile for each score in the BAMSE cohort. **(A-B)** Contains the familial, perinatal and environmental risk score in the BAMSE birth cohort combined with the genetic risk score (GRS) from the TAGC consortium (A), and the GRS from SHARE consortium **(B)**. Scores are all based on asthma ever at age 8 years, in which asthma is defined as having one or more of the following two criteria; (1) having ≥1 more attacks of wheeze in the last 12 months, (2) prescription of inhalation steroids for respiratory or lung problems prescribed by a doctor in last 12 months.

346 **Supplemental Tables**

347 **TABLE E1. Familial, perinatal and environmental weighted risk scores from 0 to 10 in PIAMA.**

Predictor	Beta	OR (95% CI)	*P* value	Score
Familial risk score				
Parental allergy*	0.32	1.38 (1.08-1.76)	0.01	1.5
Parents allergic to pets	0.36	1.43 (1.10-1.85)	0.007	1.5
Parental inhaled medication	0.43	1.54 (1.17-2.04)	0.002	2
Siblings with asthma	0.90	2.46 (1.52-3.99)	$2.55x10^{-4}$	4
Low parental education†	0.29	1.33 (1.07-1.65)	0.01	1

348

Predictor	Beta	OR (95% CI)	*P* value	Score
Perinatal risk score				
Male gender	0.37	1.44 (1.20-1.74)	$1.16x10^{-4}$	3
Breastfeeding <16 weeks	0.28	1.32 (1.09-1.60)	0.005	2
Low birth weight <2500 g	0.76	2.15 (1.24-3.70)	0.006	5

349

Predictor	Beta	OR (95% CI)	*P* value	Score
Environmental risk score				
Pets at home during pregnancy	0.19	1.21 (1.00-1.46)	0.05	2.5
Smoking mother during pregnancy‡	0.37	1.45 (1.13-1.88)	0.004	5
Older siblings living at home	0.18	1.20 (1.00-1.45)	0.05	2.5
AUC of combined risk scores		0.65		

350

351 OR, Odds ratio; CI, Confidence Interval; AUC, Area Under the Curve

352 Area Under the Curve (Receiver Operating Characteristics curve) from predicted probability of familial, perinatal and environmental risk scores.

353 * Parental allergy is defined as parental asthma ever and/or current house dust (mite) allergy and/or pet allergy and/or hay fever'

354 † Parental education is defined as an education less than the level of a bachelor's/master's degree (HBO/University in Dutch system) for at

355 least 1 of the parents.

356 ‡ Smoking mother during pregnancy is defined as smoking at least the first 4 weeks of pregnancy.

TABLE E2. Results of association with asthma ever at age 8 years per familial, perinatal, environmental and genetic weighted risk scores in PIAMA and BAMSE.

	PIAMA		BAMSE	
Risk score	**OR (95% CI)**	***P* value**	**OR (95% CI)**	***P* value**
Familial risk score	1.25 (1.19 - 1.32)	$3.17x10^{-19}$	1.46 (1.32 - 1.63)	$8.45x10^{-13}$
Perinatal risk score	1.14 (1.09 - 1.20)	$2.60x10^{-8}$	1.27 (1.14 - 1.42)	$1.00x10^{-5}$
Environmental risk score	1.08 (1.04 - 1.12)	$5.30x10^{-5}$	1.06 (0.98 - 1.15)	0.14
TAGC GRS (threshold $P<5x10^{-8}$)	1.23 (1.10 - 1.38)	$4.32x10^{-4}$	1.52 (1.19-1.96)	0.01
SHARE GRS (threshold $P<3x10^{-8}$)	1.72 (1.21 - 2.44)	0.003	4.75 (2.25-10.01)	$4.20x10^{-5}$

OR, Odds ratio; CI, Confidence Interval; GRS, Genetic risk score

Asthma is defined as having one or more of the following three criteria; (1) having ≥1 more attacks of wheeze in the last 12 months, (2) ≥1 events of shortness of breath (dyspnea) in the last 12 months, (3) prescription of inhalation steroids for respiratory or lung problems prescribed by a doctor in last 12 months. In BAMSE we used an adjusted asthma definition in which asthma was defined as having one or more of the following two criteria; (1) having ≥1 more attacks of wheeze in the last 12 months, (2) prescription of inhalation steroids for respiratory or lung problems prescribed by a doctor in last 12 months.